

EFFICIENT CALCULATION OF RISK MEASURES BY IMPORTANCE SAMPLING – THE HEAVY TAILED CASE

HENRIK HULT AND JENS SVENSSON

ABSTRACT. Computation of extreme quantiles and tail-based risk measures using standard Monte Carlo simulation can be inefficient. A method to speed up computations is provided by importance sampling. We show that importance sampling algorithms, designed for efficient tail probability estimation, can significantly improve Monte Carlo estimators of tail-based risk measures. In the heavy-tailed setting, when the random variable of interest has a regularly varying distribution, we provide sufficient conditions for the asymptotic relative error of importance sampling estimators of risk measures, such as Value-at-Risk and expected shortfall, to be small. The results are illustrated by some numerical examples.

1. INTRODUCTION

Risk measures are frequently used to quantify uncertainty in a financial or actuarial context. Many risk measures, such as Value-at-Risk and expected shortfall, depend on the tail of the loss distribution. Exact formulas for computing risk measures are only available for simple models and an alternative is to use Monte Carlo simulation. However, standard Monte Carlo can be inefficient when the function of interest depends on the occurrence of rare events. A large number of samples may be needed for accurate computation of extreme risk measures with standard Monte Carlo, resulting in high computational cost. An alternative to reduce the computational cost without loss of accuracy is provided by importance sampling. There is a vast literature on the design of importance sampling algorithms for computing rare event probabilities. However, for computing quantiles and other risk measures the literature is not as developed. Glasserman et al. (2002) propose a method for efficient computation of quantiles of a heavy-tailed portfolio. Their method is based on efficient algorithms designed for rare event probabilities. The exceedance probability is computed for a suggested initial value of the quantile. Then the quantile estimate is updated depending on the computed probability and a search algorithm is constructed to generate subsequent and more accurate quantile estimates. We follow a more direct approach suggested by Glynn (1996) for computing quantiles and extend it to handle expected shortfall.

Let us first give a brief description of the problem. Let X be a random variable with distribution μ , distribution function (d.f.) F , and continuous density f . Consider the problem of computing its p th quantile, i.e. a number λ_p such that $P(X \geq \lambda_p) \geq 1 - p$, for some $p \in (0, 1)$. If possible the quantile is calculated by

2000 *Mathematics Subject Classification.* Primary: 65C05; Secondary: 60F05.

Key words and phrases. Monte Carlo simulation; rare events; importance sampling; risk measures.

inverting the d.f.

$$\lambda_p = F^{\leftarrow}(p) := \inf\{x : 1 - F(x) \geq 1 - p\}.$$

When this is impossible an alternative is to use simulation. Computation of $F^{\leftarrow}(p)$ by standard Monte Carlo can be implemented as follows. Generate N independent copies of X , denoted X_1, \dots, X_N . The empirical distribution function (e.d.f.) of the sample is given by

$$F_N(x) = \frac{1}{N} \sum_{i=1}^N I\{X_i \leq x\}$$

and the quantile estimate is given by $F_N^{\leftarrow}(p)$. For extreme quantiles, when $1 - p$ is very small, standard Monte Carlo can be inefficient. Since only a small fraction of the sample will be located in the tail, large samples are needed to obtain reliable estimates. A rough approach for quantifying the efficiency of Monte Carlo estimates is to first consider a central limit theorem for $F_N^{\leftarrow}(p)$. Suppose (this is true under suitable conditions)

$$\sqrt{N}(F_N^{\leftarrow}(p) - F^{\leftarrow}(p)) \xrightarrow{w} N\left(0, \frac{p(1-p)}{f(F^{\leftarrow}(p))^2}\right),$$

where \xrightarrow{w} denotes weak convergence. It is desirable to have the asymptotic standard deviation of roughly the same size as the quantity $F^{\leftarrow}(p)$ we are estimating. For standard Monte Carlo the asymptotic standard deviation is

$$\frac{\sqrt{p(1-p)}}{f(F^{\leftarrow}(p))} \approx \frac{\sqrt{\bar{F}(F^{\leftarrow}(p))}}{f(F^{\leftarrow}(p))}$$

which is typically much larger than $F^{\leftarrow}(p)$ for p close to 1.

Consider, as an alternative to standard Monte Carlo, the method of importance sampling. Then the sample X_1, \dots, X_N is generated from the sampling distribution ν and the importance sampling tail e.d.f. is given by

$$\bar{F}_{\nu,N}(x) = \frac{1}{N} \sum_{i=1}^N \frac{d\mu}{d\nu}(X_i) I\{X_i > x\}.$$

The quantile estimate is then given by $(1 - \bar{F}_{\nu,N})^{\leftarrow}(p) = \inf\{x : \bar{F}_{\nu,N}(x) \leq 1 - p\}$. The goal is to choose ν to get many samples in the tail of the original distribution and with small Radon-Nikodym weights. Again, a rough evaluation of the performance may be done by studying the limiting variance σ_p^2 in a central limit theorem of the form

$$\sqrt{N}((1 - \bar{F}_{\nu,N})^{\leftarrow}(p) - F^{\leftarrow}(p)) \xrightarrow{w} N\left(0, \sigma_p^2\right).$$

It turns out that the asymptotic properties (as $p \rightarrow 1$) of σ_p^2 is closely related to the asymptotics of the second moment of importance sampling algorithms designed for computing rare event probabilities. This indicates that efficient algorithms for computing rare event probabilities are indeed useful for computing quantiles.

We use a similar approach to evaluate the performance of importance sampling algorithms for computing expected shortfall. For a random variable X with d.f. F

expected shortfall at level $p \in (0, 1)$ can be expressed as

$$\text{Expected shortfall}_p(X) = \frac{1}{1-p} \int_p^1 F^{\leftarrow}(u) du =: \gamma_p(F^{\leftarrow}).$$

The standard Monte Carlo estimate based on a sample X_1, \dots, X_N is given by $\gamma_p(F_N^{\leftarrow})$ whereas the importance sampling estimate is given by $\gamma_p((1 - \bar{F}_{\nu, N})^{\leftarrow})$. A central limit theorem is derived for the expected shortfall estimate based on importance sampling and the properties of the limiting variance are studied as $p \rightarrow 1$.

When evaluating the asymptotic variance for p close to 1 we restrict attention to the heavy-tailed case. More precisely, it is assumed that the original distribution has a regularly varying tail. This is motivated by applications, e.g. finance and insurance, where heavy-tailed data are frequently observed and evaluation of risk measures is important for risk control.

For computation of rare event tail probabilities, the importance sampling estimate of $P(X > \lambda)$ is given by $\hat{p}_\lambda = \bar{F}_{\nu, N}(\lambda)$. Typically, the performance of a rare event simulation algorithm is evaluated in terms of the relative error,

$$\text{Relative Error} = \frac{\sqrt{\text{Var}(\hat{p}_\lambda)}}{p\lambda}$$

An algorithm is said to be *asymptotically optimal* if the relative error tends to 0 as $\lambda \rightarrow \infty$. If the relative error remains bounded as $\lambda \rightarrow \infty$, the algorithm is said to have *bounded relative error*. In the heavy-tailed setting there exist several algorithms for the case where X is given by the value at time n of a random walk. Bassamboo et al. (2007) show that for such algorithms a necessary condition for them to achieve asymptotic optimality is that they are state-dependent. Dupuis et al. (2007) develop the first such algorithm, which almost achieves asymptotic optimality for regularly varying distributions. Blanchet and Glynn (2008) propose a state-dependent algorithm with bounded relative error for a more general class of heavy-tailed distributions. Blanchet and Liu (2008) consider the case where the number of steps of the random walk and λ tends to infinity simultaneously and develop an algorithm with bounded relative error. Hult and Svensson (2009) consider algorithms of the same kind as Dupuis et al. (2007), and show that they can be made asymptotically optimal.

The paper is organized as follows. In Section 2 we review some standard results from empirical process theory. In Section 3 we derive central limit theorems for empirical quantiles for the empirical measures obtained from an importance sampling algorithm. In Section 4 we consider computation of risk measures for heavy-tailed (regularly varying) random variables. Sufficient conditions for importance sampling algorithms designed for rare event probability estimation to have small asymptotic variance are provided. Finally, in Section 5 the procedure is illustrated for computation of risk measures when the variable of interest is the position of a finite random walk with regularly varying steps.

2. EMPIRICAL PROCESSES

In this section we review some basic results from the theory of empirical processes. We refer to van der Vaart and Wellner (1996) for a thorough introduction (see also Csörgő et al. (1986)).

Let $\{X_i\}_{i=1}^\infty$ be independent identically distributed random variables with distribution μ . Denote by μ_N the empirical measure based on the first N observations;

$$\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i},$$

where δ_x is the Dirac measure at x . For a collection \mathcal{F} of real valued measurable functions, the empirical measure induces a map from $\mathcal{F} \rightarrow \mathbf{R}$ by $f \mapsto \mu_N(f) = \int f d\mu_N$. Assuming $\sup_{f \in \mathcal{F}} |f(x) - \mu(f)| < \infty$ for each x , the empirical process ξ_N , given by

$$\xi_N(f) = \sqrt{N}(\mu_N(f) - \mu(f)),$$

can be viewed as a map into $l^\infty(\mathcal{F})$; the space of bounded functions $\mathcal{F} \rightarrow \mathbf{R}$ equipped with the uniform metric. The collection \mathcal{F} is called μ -Donsker if

$$\xi_N \xrightarrow{w} \xi, \quad \text{in } l^\infty(\mathcal{F})$$

and the limit is a tight Borel measurable element in $l^\infty(\mathcal{F})$.

The classical result by Donsker states that the collection of indicator functions $x \mapsto I\{x \leq t\}$ is μ -Donsker with the limiting process $B \circ \mu$, where B is a Brownian bridge on $[0, 1]$ (see van der Vaart and Wellner (1996), pp. 81-82). In this paper we will be particularly concerned with the collection $\mathcal{F}_{a,b}$ of indicator functions $x \mapsto I\{a < x \leq t\}$ for $-\infty \leq a < t \leq b \leq \infty$, which also is μ -Donsker for any probability distribution μ . To simplify notation we will often write $\xi_N \xrightarrow{w} \xi$ in $l^\infty[a, b]$ for $\xi_N \xrightarrow{w} \xi$ in $l^\infty(\mathcal{F}_{a,b})$.

To obtain convergence results for mappings of the empirical process it is useful to apply the functional delta-method. Let \mathbf{E}_1 and \mathbf{E}_2 be two metric spaces. A mapping $\phi : \mathbf{E}_1 \rightarrow \mathbf{E}_2$ is said to be *Hadamard differentiable at θ tangentially to $E_0 \subset \mathbf{E}_1$* if there is a continuous mapping $\phi'_\theta(h) : \mathbf{E}_1 \rightarrow \mathbf{E}_2$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \rightarrow \phi'_\theta(h)$$

for all sequences $t_n \rightarrow 0$ and $h_n \rightarrow h$, where $h \in E_0$.

Theorem 2.1 (Functional delta-method, c.f. van der Vaart and Wellner (1996), Theorem 3.9.4). *Let $\phi : \mathbf{E}_1 \rightarrow \mathbf{E}_2$ be Hadamard differentiable at θ tangentially to $E_0 \subset \mathbf{E}_1$. Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables taking values in \mathbf{E}_1 . Suppose $r_n(X_n - \theta) \xrightarrow{w} X \in E_0$ for some sequence of constants $r_n \rightarrow \infty$. Then $r_n(\phi(X_n) - \phi(\theta)) \xrightarrow{w} \phi'_\theta(X)$.*

For any càdlàg function $F : \mathbf{R} \rightarrow \mathbf{R}$, define the inverse map ϕ_p by

$$\phi_p(F) = F^\leftarrow(p) = \inf\{u : F(u) \geq p\}, \quad p \in (0, 1).$$

The following result shows that the functional delta-method implies the convergence of quantiles.

Proposition 2.2 (c.f. Lemma 3.9.23 and Example 3.9.24 in van der Vaart and Wellner (1996)). *Let $\{X_i\}_{i=1}^\infty$ be a sequence of independent and identically distributed random variables with d.f. F . Suppose F has a continuous density $f > 0$ with respect to the Lebesgue measure on the interval $[F^\leftarrow(p) - \epsilon, F^\leftarrow(q) + \epsilon]$, for $0 < p < q < 1$ and $\epsilon > 0$. Then*

$$\sqrt{N}(F_N^\leftarrow - F^\leftarrow) \xrightarrow{w} \frac{B}{f \circ F^\leftarrow}, \quad \text{in } l^\infty[p, q],$$

where the right-hand side refers to the random function $u \mapsto \frac{B(u)}{f(F^{-1}(u))}$.

3. EMPIRICAL PROCESSES AND IMPORTANCE SAMPLING

The empirical measure resulting from a random sample of an importance sampling algorithm with sampling distribution ν can be used to approximate important parts of the original distribution. In our context ν is chosen to give a good approximation of the extreme tail of the original distribution.

Let $\{X_i\}_{i=1}^\infty$ be independent identically distributed with distribution ν . The empirical measure with likelihood ratio weights and the corresponding tail empirical distribution are written

$$\begin{aligned}\mu_{N,\nu} &= \frac{1}{N} \sum_{i=1}^N w(X_i) \delta_{X_i} = \frac{1}{N} \sum_{i=1}^N \frac{d\mu}{d\nu}(X_i) \delta_{X_i}, \\ \bar{F}_{\nu,N}(t) &= \mu_{N,\nu}(I\{\cdot > t\}).\end{aligned}\tag{3.1}$$

Let \mathcal{F}_a be the collection of indicator functions $I\{\cdot > t\}$ with $t \geq a$. For the importance sampling estimators we are concerned with a central limit theorem of the form

$$\sqrt{N}(\mu_{N,\nu} - \mu) \xrightarrow{w} Z \quad \text{in } l^\infty(\mathcal{F}_a),$$

which we write, with slight abuse of notation, as

$$\sqrt{N}(\bar{F}_{\nu,N} - \bar{F}) \xrightarrow{w} Z, \quad \text{in } l^\infty[a, \infty].$$

Note that $\mu_{N,\nu}(f) = \nu_N(wf)$ and $\mu(f) = \nu(wf)$, where $w = d\mu/d\nu$. Therefore the central limit theorem can be stated by saying that the collection $w\mathcal{F}_a = \{wf : f \in \mathcal{F}_a\}$ is ν -Donsker. By the permanence properties of Donsker classes (see van der Vaart and Wellner (1996) Section 2.10) this follows when \mathcal{F}_a is ν -Donsker and $E_\nu w(X)^2 I\{X > a\} < \infty$.

To identify the limiting process Z we first need to calculate the covariance function of the process $\bar{F}_{\nu,N}$.

Lemma 3.1. *Let $\{X_i\}_{i=1}^\infty$ be independent and identically distributed with distribution ν with $\mu \ll \nu$ and $w = d\mu/d\nu$. If $Ew(X_1)^2 I\{X_1 > a\} < \infty$, for some $a \geq -\infty$, then, for $y \geq x > a$,*

$$\varrho(x, y) := N \text{Cov}(\bar{F}_{\nu,N}(x), \bar{F}_{\nu,N}(y)) = E_\nu w(X_1)^2 I\{X_1 > y\} - \bar{F}(x)\bar{F}(y). \tag{3.2}$$

Proof. This is a direct calculation. Indeed,

$$\begin{aligned}
& \text{Cov}(\bar{F}_{\nu,N}(x), \bar{F}_{\nu,N}(y)) \\
&= E_\nu(\bar{F}_{\nu,N}(x)\bar{F}_{\nu,N}(y)) - E_\nu(\bar{F}_{\nu,N}(x))E_\nu(\bar{F}_{\nu,N}(y)) \\
&= \frac{1}{N^2} \sum_{i,j} E_\nu(w(X_i)w(X_j)I\{X_i > x\}I\{X_j > y\}) - \bar{F}(x)\bar{F}(y) \\
&= \frac{1}{N} E_\nu(w(X_1)^2 I\{X_1 > x\}I\{X_1 > y\}) \\
&\quad + \frac{1}{N^2} \sum_{i \neq j} E_\nu(w(X_i)w(X_j)I\{X_i > x\}I\{X_j > y\}) - \bar{F}(x)\bar{F}(y) \\
&= \frac{1}{N} E_\nu(w(X_1)^2 I\{X_1 > y\}) \\
&\quad + \frac{N^2 - N}{N^2} E_\nu(w(X_i)w(X_j)I\{X_i > x\}I\{X_j > y\}) - \bar{F}(x)\bar{F}(y) \\
&= \frac{1}{N} E_\nu(w(X_1)^2 I\{X_1 > y\}) + \frac{N^2 - N}{N^2} \bar{F}(x)\bar{F}(y) - \bar{F}(x)\bar{F}(y) \\
&= \frac{1}{N} E_\nu(w(X_1)^2 I\{X_1 > y\}) - \frac{1}{N} \bar{F}(x)\bar{F}(y).
\end{aligned}$$

□

Note that if $w = 1$, i.e. the sampling measure is the original measure, then the covariance function becomes $\bar{F}(y) - \bar{F}(y)\bar{F}(x) = \bar{F}(y)(1 - \bar{F}(x)) = F(x)(1 - F(y))$, $y > x$, which corresponds to a Brownian bridge evaluated at F .

Now we are ready to state the central limit theorem for the tail empirical distribution of an importance sampling algorithm.

Proposition 3.2. *Let Z be a centered Gaussian process with covariance function ϱ given by (3.2). If $E_\nu w(X_1)^2 I\{X_1 > a\} < \infty$ for some $a \geq -\infty$, then*

$$\sqrt{N}(\bar{F}_{\nu,N} - \bar{F}) \xrightarrow{w} Z, \quad \text{in } l^\infty[a, \infty].$$

Proof. We have already seen that $E_\nu w(X_1)^2 I\{X_1 > a\} < \infty$ implies that \mathcal{F}_a is ν -Donsker. Hence, we need only to identify the limiting process Z . Denote $\xi_N(x) = \sqrt{N}(\bar{F}_{\nu,N}(x) - F(x))$. By the multivariate central limit theorem the finite dimensional distributions converge; for any x_1, \dots, x_k with $x_i > a$,

$$(\xi_N(x_1), \dots, \xi_N(x_k)) \xrightarrow{w} N(0, \Sigma),$$

where the entries of $\Sigma_{ij} = \varrho(x_i, x_j)$. This determines that the limiting process must be Z . □

We proceed with the asymptotic normality of the quantile transform. The proof is very similar to that of Proposition 2.2 and therefore omitted.

Proposition 3.3. *Let Z be a centered Gaussian process with covariance function ϱ in (3.2). Suppose F has a continuous density $f > 0$ with respect to the Lebesgue measure on the interval $[F^\leftarrow(p) - \epsilon, F^\leftarrow(q) + \epsilon]$, for $0 < p < q < 1$ and $\epsilon > 0$. $E_\nu w(X_1)^2 I\{X_1 > F^\leftarrow(p) - \epsilon\} < \infty$, then*

$$\sqrt{N}((1 - \bar{F}_{\nu,N})^\leftarrow - F^\leftarrow) \xrightarrow{w} \frac{Z(F^\leftarrow)}{f(F^\leftarrow)}, \quad \text{in } l^\infty[p, q].$$

Next consider a central limit theorem for an importance sampling estimate of expected shortfall. For a non-decreasing càdlàg function F^\leftarrow and $0 < p < 1$ we use the notation

$$\gamma_p(F^\leftarrow) = \frac{1}{1-p} \int_p^1 F^\leftarrow(u) du.$$

Recall that expected shortfall at level p for a random variable X with d.f. F is given by $\gamma_p(F^\leftarrow)$ and the importance sampling estimate based on a sample X_1, \dots, X_N with sampling distribution ν is given by $\gamma_p((1 - \bar{F}_{\nu, N})^\leftarrow)$.

Proposition 3.4. *Assume the hypotheses of Proposition 3.3. If, in addition, $\int_{F^\leftarrow(p)}^\infty \int_{F^\leftarrow(p)}^\infty \varrho(x, y) dx dy < \infty$ and $\varrho(x, x) = o([f(x)/\bar{F}(x)]^2)$, as $x \rightarrow \infty$, then*

$$\sqrt{N}(\gamma_p((1 - \bar{F}_{\nu, N})^\leftarrow) - \gamma_p(F^\leftarrow)) \xrightarrow{w} \frac{1}{1-p} \int_p^1 \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du,$$

as $N \rightarrow \infty$.

Proof. Let q and ε be arbitrary with $p < q < 1$ and $\varepsilon > 0$. Since

$$\begin{aligned} & P\left(\left|\sqrt{N}[\gamma_p((1 - \bar{F}_{\nu, N})^\leftarrow) - \gamma_p(F^\leftarrow)] - \frac{1}{1-p} \int_p^1 \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du\right| > \varepsilon\right) \\ & \leq P\left(\left|\frac{\sqrt{N}}{1-p} \left[\int_p^q (1 - \bar{F}_{\nu, N})^\leftarrow(u) - \int_p^q F^\leftarrow(u) du\right] - \frac{1}{1-p} \int_p^q \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du\right| > \varepsilon/3\right) \end{aligned} \quad (3.3)$$

$$+ P\left(\left|\sqrt{N} \left|\frac{1}{1-p} \int_q^1 (1 - \bar{F}_{\nu, N})^\leftarrow(u) - F^\leftarrow(u) du\right| > \varepsilon/3\right)\right) \quad (3.4)$$

$$+ P\left(\left|\frac{1}{1-p} \int_q^1 \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du\right| > \varepsilon/3\right) \quad (3.5)$$

it is sufficient to show that each of the three terms converges to 0, as first $N \rightarrow \infty$ and then $q \rightarrow 1$.

Consider first (3.3). Let $\gamma_{p,q}$ be the map defined by

$$\gamma_{p,q}(H) = \frac{1}{1-p} \int_p^q H(u) du,$$

on the set D_γ of all non-decreasing càdlàg functions H . Since $\gamma_{p,q}$ is linear it is Hadamard differentiable on D_γ with derivative $\gamma'_{p,q}(h) = \gamma_{p,q}(h)$. In particular, Proposition 3.3 and the delta-method imply that

$$\sqrt{N}(\gamma_{p,q}((1 - \bar{F}_{\nu, N})^\leftarrow) - \gamma_{p,q}(F^\leftarrow)) \xrightarrow{w} \frac{1}{1-p} \int_p^q \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du.$$

This takes care of (3.3). Next consider (3.5). By Chebyshev's inequality

$$\begin{aligned}
& P\left(\left|\frac{1}{1-p} \int_q^1 \frac{Z(F^{\leftarrow}(u))}{f(F^{\leftarrow}(u))} du\right| > \varepsilon/3\right) \\
& \leq \left(\frac{3(1-p)}{\varepsilon}\right)^2 \text{Var}\left(\int_q^1 \frac{Z(F^{\leftarrow}(u))}{f(F^{\leftarrow}(u))} du\right) \\
& \leq \left(\frac{3(1-p)}{\varepsilon}\right)^2 \int_q^1 \int_q^1 \frac{\varrho(F^{\leftarrow}(u), F^{\leftarrow}(v))}{f(F^{\leftarrow}(u))f(F^{\leftarrow}(v))} dudv \\
& = \left(\frac{3(1-p)}{\varepsilon}\right)^2 \int_{F^{\leftarrow}(q)}^\infty \int_{F^{\leftarrow}(q)}^\infty \varrho(x, y) dx dy.
\end{aligned}$$

Since the integral is finite, this converges to 0 as $q \rightarrow 1$.

It remains to consider (3.4). First, write

$$\begin{aligned}
& \sqrt{N} \left| \frac{1}{1-p} \int_q^1 (1 - \bar{F}_{\nu, N})^{\leftarrow}(u) - F^{\leftarrow}(u) du \right| \\
& \leq \frac{\sqrt{N}}{1-p} \left| \int_{(1-\bar{F}_{\nu, N})^{\leftarrow}(q)}^\infty \bar{F}_{\nu, N}(x) dx - \int_{F^{\leftarrow}(q)}^\infty \bar{F}(x) dx \right| \\
& \quad + \frac{\sqrt{N}(1-q)}{1-p} \left| (1 - \bar{F}_{\nu, N})^{\leftarrow}(q) - F^{\leftarrow}(q) \right| \\
& \leq \frac{\sqrt{N}}{1-p} \left| \int_{F^{\leftarrow}(q)}^\infty \bar{F}_{\nu, N}(x) - \bar{F}(x) dx \right| \tag{3.6}
\end{aligned}$$

$$+ \frac{\sqrt{N}}{1-p} \left| \int_{(1-\bar{F}_{\nu, N})^{\leftarrow}(q)}^{F^{\leftarrow}(q)} \bar{F}_{\nu, N}(x) dx \right| I\{(1 - \bar{F}_{\nu, N})^{\leftarrow}(q) \leq F^{\leftarrow}(q)\} \tag{3.7}$$

$$+ \frac{\sqrt{N}}{1-p} \left| \int_{F^{\leftarrow}(q)}^{(1-\bar{F}_{\nu, N})^{\leftarrow}(q)} \bar{F}_{\nu, N}(x) dx \right| I\{(1 - \bar{F}_{\nu, N})^{\leftarrow}(q) \geq F^{\leftarrow}(q)\} \tag{3.8}$$

$$+ \frac{\sqrt{N}(1-q)}{1-p} \left| (1 - F_{\nu, N})^{\leftarrow}(q) - F^{\leftarrow}(q) \right|. \tag{3.9}$$

First consider (3.6). By Proposition 3.2 and the delta method

$$\begin{aligned}
& \lim_{N \rightarrow \infty} P\left(\frac{\sqrt{N}}{1-p} \left| \int_{F^{\leftarrow}(q)}^\infty \bar{F}_{\nu, N}(x) - \bar{F}(x) dx \right| > \varepsilon/12\right) \\
& = P\left(\frac{1}{1-p} \left| \int_{F^{-1}(q)}^\infty Z(x) dx \right| > \varepsilon/12\right) \\
& \leq \left(\frac{12}{\varepsilon(1-p)}\right)^2 \text{Var}\left(\int_{F^{\leftarrow}(q)}^\infty Z(x) dx\right) \\
& = \left(\frac{12}{\varepsilon(1-p)}\right)^2 \int_{F^{\leftarrow}(q)}^\infty \int_{F^{\leftarrow}(q)}^\infty \varrho(x, y) dx dy.
\end{aligned}$$

Since $F^{\leftarrow}(q) \rightarrow \infty$ as $q \rightarrow 1$ and the integral is finite the expression in the last display can be made arbitrarily small. Next, consider (3.7). This term is bounded

from above by

$$\begin{aligned} & \frac{\sqrt{N}}{1-p} \bar{F}_{\nu,N} \left((1 - \bar{F}_{\nu,N})^{\leftarrow}(q) \right) \left(F^{\leftarrow}(q) - (1 - \bar{F}_{\nu,N})^{\leftarrow}(q) \right) \\ &= \frac{1-q}{1-p} \sqrt{N} (F^{\leftarrow}(q) - (1 - \bar{F}_{\nu,N})^{\leftarrow}(q)), \end{aligned}$$

where we have used that $\bar{F}_{\nu,N}((1 - \bar{F}_{\nu,N})^{\leftarrow}(q)) \geq 1 - q$. By Proposition 3.3,

$$\begin{aligned} & \lim_{N \rightarrow \infty} P \left(\frac{1-q}{1-p} \sqrt{N} |F^{\leftarrow}(q) - (1 - \bar{F}_{\nu,N})^{\leftarrow}(q)| > \varepsilon/12 \right) \\ &= P \left(\frac{1-q}{1-p} \left| \frac{Z(F^{\leftarrow}(q))}{f(F^{\leftarrow}(q))} \right| > \varepsilon/12 \right) \\ &\leq \left(\frac{12}{(1-p)\varepsilon} \right)^2 (1-q)^2 \frac{\varrho(F^{\leftarrow}(q), F^{\leftarrow}(q))}{f(F^{\leftarrow}(q))^2} \\ &= \left(\frac{12}{(1-p)\varepsilon} \right)^2 \bar{F}(F^{\leftarrow}(q))^2 \frac{\varrho(F^{\leftarrow}(q), F^{\leftarrow}(q))}{f(F^{\leftarrow}(q))^2}. \end{aligned}$$

This converges to 0 as $q \rightarrow 1$ since $\varrho(x, x) = o([f(x)/\bar{F}(x)]^2)$.

Similarly, (3.8) is bounded from above by

$$\frac{\sqrt{N}}{1-p} \bar{F}_{\nu,N}(F^{\leftarrow}(q)) \left((1 - \bar{F}_{\nu,N})^{\leftarrow}(q) - F^{\leftarrow}(q) \right).$$

This can be treated just like the previous term since, by Proposition 3.3,

$$\begin{aligned} & \lim_{N \rightarrow \infty} P \left(\frac{\bar{F}_{\nu,N}(F^{\leftarrow}(q))}{1-p} \sqrt{N} |(1 - \bar{F}_{\nu,N})^{\leftarrow}(q) - F^{\leftarrow}(q)| > \varepsilon/12 \right) \\ &= P \left(\frac{1-q}{1-p} \left| \frac{Z(F^{\leftarrow}(q))}{f(F^{\leftarrow}(q))} \right| > \varepsilon/12 \right). \end{aligned}$$

Finally, (3.9) can be treated the same way since, by Proposition 3.3,

$$\begin{aligned} & \lim_{N \rightarrow \infty} P \left(\frac{\sqrt{N}(1-q)}{1-p} \left| (1 - F_{\nu,N})^{\leftarrow}(q) - F^{\leftarrow}(q) \right| > \varepsilon/12 \right) \\ &= P \left(\frac{1-q}{1-p} \left| \frac{Z(F^{\leftarrow}(q))}{f(F^{\leftarrow}(q))} \right| > \varepsilon/12 \right). \end{aligned}$$

This completes the proof. \square

4. EFFICIENT CALCULATION OF RISK MEASURES IN THE HEAVY-TAILED SETTING

In the previous section we established central limit theorems for importance sampling estimates of Value-at-Risk (i.e. quantiles) and expected shortfall. In this section we study the limiting variance of the central limit theorems as a function of p when p is close to 1. The main requirement is that the asymptotic standard deviation coming from the central limit theorem is roughly of the same size as the quantity we are trying to compute, when p is close to 1. We only consider the case when the original distribution is heavy-tailed, in the sense that \bar{F} is regularly varying.

There are three main assumptions in this section.

- We assume that the original distribution of interest has a regularly varying tail. That is, there exists $\alpha > 0$ such that

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}, \quad x > 0. \quad (4.1)$$

- We assume that there is an available explicit asymptotic approximation for $\overline{F}(x)$. More precisely, we know a non-increasing function U such that $U \sim \overline{F}$, i.e.

$$\lim_{t \rightarrow \infty} \frac{U(t)}{\overline{F}(t)} = 1. \quad (4.2)$$

- We assume that we can construct sampling measures ν_λ with bounded relative error for computing rare event probabilities of the type $\overline{F}(\lambda) = P(X > \lambda)$; i.e.

$$\limsup_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X_1)^2 I\{X_1 > \lambda\}}{\overline{F}(\lambda)^2} < \infty. \quad (4.3)$$

4.1. Computation of quantiles – Value-at-Risk. For a random variable X with d.f. F the Value-at-Risk at level p is defined as the p th quantile; $\text{VaR}_p(X) = F^\leftarrow(p)$. Given $p \in (0, 1)$ close to 1, the importance sampling estimate based on independent and identically distributed samples with sampling distribution ν is given by $(1 - \overline{F}_{\nu, N})^\leftarrow(p)$. Then Proposition 3.3 and Lemma 3.1 determines the asymptotic variance as

$$\begin{aligned} \text{Var} \left(\frac{Z(F^\leftarrow(p))}{f(F^\leftarrow(p))} \right) &= \frac{\varrho(F^\leftarrow(p), F^\leftarrow(p))}{f(F^\leftarrow(p))^2} \\ &= \frac{E_\nu w(X_1)^2 I\{X_1 > F^\leftarrow(p)\} - \overline{F}(F^\leftarrow(p))^2}{f(F^\leftarrow(p))^2} \\ &= \frac{E_\nu w(X_1)^2 I\{X_1 > F^\leftarrow(p)\} - (1-p)^2}{f(F^\leftarrow(p))^2} := \sigma_p^2. \end{aligned}$$

To control the asymptotic variance it seems like a good choice to use an efficient rare event simulation algorithm designed for efficient computation of $P(X > F^\leftarrow(p))$. That is, with sampling distribution $\nu_{F^\leftarrow(p)}$. This is of course impossible since $F^\leftarrow(p)$ is unknown. However, the asymptotic approximation U may be helpful. Note that since U is monotone it has an inverse U^\leftarrow and by regular variation $(1 - U)^\leftarrow \sim F^\leftarrow$ as $p \rightarrow 1$. Thus, it seems reasonable to use ν_{u_p} where $u_p = (1 - U)^\leftarrow(p)$. This is justified by the next result.

Proposition 4.1. *Suppose (4.1)-(4.3) hold. If there exists $c_0 < 1$ such that*

$$\limsup_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X_1)^2 I\{X_1 > c_0 \lambda\}}{\overline{F}(\lambda)^2} < \infty, \quad (4.4)$$

then the sampling measures ν_{u_p} satisfy

$$\limsup_{p \rightarrow 1} \frac{\sigma_p^2}{F^\leftarrow(p)^2} < \infty.$$

Proof. First note that (4.4) implies that

$$\limsup_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X_1)^2 I\{X_1 > c\lambda\}}{\overline{F}(\lambda)^2} < \infty,$$

for any $c \geq c_0$. Take $\varepsilon \in (0, 1 - c_0)$. Then there exists p_0 such that $\frac{F^\leftarrow(p)}{u_p} > 1 - \varepsilon$, for each $p \geq p_0$. In particular,

$$\begin{aligned} \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(p)\}}{\bar{F}(u_p)^2} \\ \leq \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > (1 - \varepsilon)u_p\}}{\bar{F}(u_p)^2} < \infty \end{aligned} \quad (4.5)$$

Since $u_p \sim F^\leftarrow(p)$ and $xf(x) \sim \alpha \bar{F}(x)$, by Karamata's theorem, it follows that

$$\lim_{p \rightarrow 1} \frac{\bar{F}(U^\leftarrow(p))^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} = \lim_{p \rightarrow 1} \frac{\bar{F}(F^\leftarrow(p))^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} = \frac{1}{\alpha^2}, \quad (4.6)$$

$$\lim_{p \rightarrow 1} \frac{(1 - p)^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} = \lim_{p \rightarrow 1} \frac{\bar{F}(F^\leftarrow(p))^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} = \frac{1}{\alpha^2}. \quad (4.7)$$

By (4.5), (4.6), and (4.7) it follows that

$$\begin{aligned} \limsup_{p \rightarrow 1} \frac{\sigma_p^2}{F^\leftarrow(p)^2} &= \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(p)\} - (1 - p)^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} \\ &= \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(p)\}}{\bar{F}(U^\leftarrow(p))^2} \frac{\bar{F}(U^\leftarrow(p))^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} \\ &\quad - \frac{(1 - p)^2}{F^\leftarrow(p)^2 f(F^\leftarrow(p))^2} < \infty. \end{aligned}$$

□

Under somewhat stronger assumptions it is possible to reach a more explicit asymptotic bound for $\frac{\sigma_p^2}{F^\leftarrow(p)^2}$.

Proposition 4.2. *Suppose (4.1) and (4.2) hold. Suppose additionally that there exist $c_0 < 1$ and a function φ , continuous at 1, such that, for $c \geq c_0$,*

$$\lim_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X_1)^2 I\{X_1 > c\lambda\}}{\bar{F}(\lambda)^2} \leq \varphi(c). \quad (4.8)$$

Then the sampling measures ν_{u_p} satisfy

$$\lim_{p \rightarrow 1} \frac{\sigma_p^2}{F^\leftarrow(p)^2} \leq \frac{\varphi(1) - 1}{\alpha^2}.$$

Remark 4.3. If the asymptotic quantile approximation based on U always underestimates the true quantile, i.e. $u_p \leq F^\leftarrow(p)$ for each p , then one can take $c_0 = 1$ in Proposition 4.1 and Proposition 4.2.

Proof. First note that $(1 - U)^\leftarrow \sim F^\leftarrow$. Take $\varepsilon \in (0, 1 - c_0)$. Then there exists p_0 such that $\frac{F^\leftarrow(p)}{u_p} > 1 - \varepsilon$, for each $p \geq p_0$. In particular,

$$\limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(p)\}}{\bar{F}(u_p)} \leq \varphi(1 - \varepsilon).$$

Since $\varepsilon > 0$ is arbitrary and φ continuous at 1 it is possible to let $\varepsilon \rightarrow 0$ and get the upper bound

$$\limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^{\leftarrow}(p)\}}{\overline{F}(u_p)} \leq \varphi(1).$$

Then, by (4.6), and (4.7) it follows that

$$\begin{aligned} \limsup_{p \rightarrow 1} \frac{\sigma_p^2}{F^{\leftarrow}(p)^2} &= \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^{\leftarrow}(p)\} - (1-p)^2}{F^{\leftarrow}(p)^2 f(F^{\leftarrow}(p))^2} \\ &= \limsup_{p \rightarrow 1} \frac{E_{\nu_{u_p}} w_{u_p}(X_1)^2 I\{X_1 > F^{\leftarrow}(p)\}}{\overline{F}(U^{\leftarrow}(p))^2} \frac{\overline{F}(U^{\leftarrow}(p))^2}{F^{\leftarrow}(p)^2 f(F^{\leftarrow}(p))^2} \\ &\quad - \frac{(1-p)^2}{F^{\leftarrow}(p)^2 f(F^{\leftarrow}(p))^2} = \frac{\varphi(1) - 1}{\alpha^2}. \end{aligned}$$

□

4.2. Expected Shortfall. Next we consider the properties, when p is close to 1, of the asymptotic variance in the central limit theorem, Proposition 3.4, for expected shortfall.

Proposition 4.4. *Let $\alpha > 2$. Suppose (4.1) and (4.2) hold. Suppose additionally that there exist $c_0 < 1$ and a non-increasing function φ , regularly varying with index $-\alpha$, such that, for $c \geq c_0$,*

$$\limsup_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X_1)^2 I\{X_1 > c\lambda\}}{\overline{F}(\lambda)^2} \leq \varphi(c). \quad (4.9)$$

Then the sampling measures ν_{u_p} satisfy

$$\limsup_{p \rightarrow 1} \frac{\text{Var} \left(\frac{1}{1-p} \int_p^1 \frac{Z(F^{\leftarrow}(u))}{f(F^{\leftarrow}(u))} du \right)}{\gamma_p(F^{\leftarrow})^2} < \infty.$$

Moreover, if (4.9) holds with $h(c) = Kc^{-\alpha}$, for some constant $K \in (0, \infty)$, then

$$\limsup_{p \rightarrow 1} \frac{\text{Var} \left(\frac{1}{1-p} \int_p^1 \frac{Z(F^{\leftarrow}(u))}{f(F^{\leftarrow}(u))} du \right)}{\gamma_p(F^{\leftarrow})^2} \leq \frac{1}{\alpha^2} \left(\frac{2K(\alpha-1)}{\alpha-2} - 1 \right).$$

Proof. By Lemma 3.1,

$$\begin{aligned} \varrho(F^{\leftarrow}(u), F^{\leftarrow}(v)) &= E_{\nu_{u_p}} \left(w_{u_p}(X_1)^2 I\{X_1 > F^{\leftarrow}(u)\} I\{X_1 > F^{\leftarrow}(v)\} \right) - \overline{F}(F^{\leftarrow}(u)) \overline{F}(F^{\leftarrow}(v)) \\ &= E_{\nu_{u_p}} \left(w_{u_p}(X_1)^2 I\{X_1 > F^{\leftarrow}(u)\} I\{X_1 > F^{\leftarrow}(v)\} \right) - (1-u)(1-v), \end{aligned}$$

which implies, by Proposition 3.2, that

$$\begin{aligned}
& \frac{\text{Var}\left(\frac{1}{1-p} \int_p^1 \frac{Z(F^\leftarrow(q))}{f(F^\leftarrow(q))} dq\right)}{\gamma_p(F^\leftarrow)^2} \\
&= \frac{1}{(1-p)^2 \gamma_p(F^\leftarrow)^2} \int_p^1 \int_p^1 \frac{E(Z(F^\leftarrow(u))Z(F^\leftarrow(v)))}{f(F^\leftarrow(u))f(F^\leftarrow(v))} dudv \\
&= \frac{1}{(1-p)^2 \gamma_p(F^\leftarrow)^2} \int_p^1 \int_p^1 \frac{\varrho(F^\leftarrow(u), F^\leftarrow(v))}{f(F^\leftarrow(u))f(F^\leftarrow(v))} dudv \\
&= \frac{1}{(1-p)^2 \gamma_p(F^\leftarrow)^2} \\
&\quad \times \int_p^1 \int_p^1 \frac{E_{\nu_{u_p}}\left(w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(u)\} I\{X_1 > F^\leftarrow(v)\}\right)}{f(F^\leftarrow(u))f(F^\leftarrow(v))} dudv \quad (4.10)
\end{aligned}$$

$$- \frac{1}{(1-p)^2 \gamma_p(F^\leftarrow)^2} \int_p^1 \int_p^1 \frac{(1-u)(1-v)}{f(F^\leftarrow(u))f(F^\leftarrow(v))} dudv. \quad (4.11)$$

Consider first (4.11). By Karamata's theorem $f(F^\leftarrow(u)) \sim \frac{\alpha}{F^\leftarrow(u)} \overline{F}(F^\leftarrow(u)) = \frac{\alpha(1-u)}{F^\leftarrow(u)}$, as $u \rightarrow 1$, and therefore

$$\lim_{p \rightarrow 1} \frac{\frac{1}{(1-p)} \int_p^1 \frac{(1-u)}{f(F^\leftarrow(u))} du}{\gamma_p(F^\leftarrow)} = \lim_{p \rightarrow 1} \frac{\int_p^1 \alpha^{-1} F^\leftarrow(u) du}{\int_p^1 F^\leftarrow(u) du} = \alpha^{-1}.$$

This yields,

$$\lim_{p \rightarrow 1} \frac{\frac{1}{(1-p)^2} \int_p^1 \int_p^1 \frac{(1-u)(1-v)}{f(F^\leftarrow(u))f(F^\leftarrow(v))} dudv}{\gamma_p(F^\leftarrow)^2} = \alpha^{-2}.$$

Next rewrite (4.10) for $v > u$ as

$$\begin{aligned}
& \frac{2}{(1-p)^2 \gamma_p(F^\leftarrow)^2} \\
& \times \int_p^1 \frac{1}{f(F^\leftarrow(u))} \left[\int_u^1 \frac{E_{\nu_{u_p}}\left(w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(v)\}\right)}{f(F^\leftarrow(v))} dv \right] du
\end{aligned}$$

Then, the inner integrand can be written

$$\begin{aligned}
& \frac{E_{\nu_{u_p}}\left(w_{u_p}(X_1)^2 I\{X_1 > F^\leftarrow(v)\}\right)}{f(F^\leftarrow(v))} \\
&= \frac{E_{\nu_{u_p}}\left(w_{u_p}(X_1)^2 I\{X_1 > \frac{F^\leftarrow(v)}{u_p} u_p\}\right)}{\overline{F}(u_p)^2} \frac{\overline{F}(u_p)^2}{\overline{F}(F^\leftarrow(p))^2} \frac{\overline{F}(F^\leftarrow(p))^2}{f(F^\leftarrow(v))} \\
&\lesssim \varphi\left(\frac{F^\leftarrow(v)}{u_p}\right) \frac{(1-p)^2}{f(F^\leftarrow(v))}, \quad (4.12)
\end{aligned}$$

where we have used (4.9). By Potter's bound there exists, for each $\varepsilon > 0$, a constant C_ε such that $\varphi(F^\leftarrow(v)/u_p) \leq C_\varepsilon (F^\leftarrow(v)/u_p)^{-\alpha+\varepsilon}$. Take $0 < \varepsilon < 2 - \alpha$.

The asymptotics of the integral in (4.10) can now be determined as

$$\begin{aligned}
& \int_p^1 \frac{1}{f(F^\leftarrow(u))} \left[\int_u^1 \frac{E_{\nu_{up}}(w_{up}(X_1)^2 I\{X_1 > F^\leftarrow(v)\})}{f(F^\leftarrow(v))} dv \right] du \\
& \lesssim C_\varepsilon u_p^{\alpha-\varepsilon} (1-p)^2 \int_p^1 \frac{1}{f(F^\leftarrow(u))} \int_u^1 \frac{F^\leftarrow(v)^{-\alpha+\varepsilon}}{f(F^\leftarrow(v))} dv \\
& = C_\varepsilon u_p^{\alpha-\varepsilon} (1-p)^2 \int_p^1 \frac{1}{f(F^\leftarrow(u))} \int_{F^\leftarrow(u)}^\infty y^{-\alpha+\varepsilon} dy du \\
& = C_\varepsilon u_p^{\alpha-\varepsilon} (1-p)^2 \int_{F^\leftarrow(p)}^\infty \frac{x^{1-\alpha+\varepsilon}}{\alpha-\varepsilon-1} dx \\
& = C_\varepsilon u_p^{\alpha-\varepsilon} (1-p)^2 \frac{F^\leftarrow(p)^{2-\alpha+\varepsilon}}{(\alpha-\varepsilon-1)(\alpha-\varepsilon-2)} \\
& \sim C_\varepsilon (1-p)^2 \frac{F^\leftarrow(p)^2}{(\alpha-\varepsilon-1)(\alpha-\varepsilon-2)}.
\end{aligned}$$

By Karamata's theorem

$$\begin{aligned}
\gamma_p(F^\leftarrow) &= \frac{1}{1-p} \int_p^1 F^\leftarrow(u) du = \frac{1}{1-p} \int_{F^\leftarrow(p)}^\infty x f(x) dx \\
&\sim \frac{1}{1-p} \frac{\alpha}{\alpha-1} F^\leftarrow(p) \overline{F}(F^\leftarrow(p)) \\
&= \frac{\alpha}{\alpha-1} F^\leftarrow(p).
\end{aligned}$$

Putting everything together, the expression in (4.10) is asymptotically bounded.

Moreover, if (4.9) holds with $\varphi(c) = Kc^{-\alpha}$, for some constant $K \in (0, \infty)$, then it is possible to take $C_\varepsilon = K$ and $\varepsilon = 0$. This results in

$$\limsup_{p \rightarrow 1} \frac{\text{Var} \left(\frac{1}{1-p} \int_p^1 \frac{Z(F^\leftarrow(u))}{f(F^\leftarrow(u))} du \right)}{\gamma_p(F^\leftarrow)^2} \leq \frac{1}{\alpha^2} \left(\frac{2K(\alpha-1)}{\alpha-2} - 1 \right).$$

□

5. EXAMPLES AND NUMERICAL ILLUSTRATIONS

In this section we use the methods presented in the previous sections to design efficient algorithms for computing Value-at-Risk and expected shortfall of a random variable X which is the value at time $n \geq 1$ of a heavy-tailed random walk. More precisely,

$$X = \sum_{i=1}^n Z_i, \tag{5.1}$$

where Z_i , $i = 1, \dots, n$, are i.i.d. and regularly varying with tail index α . We will use F_X and F_Z to denote the d.f. of X and Z_1 , respectively. We write f_X and f_Z for the corresponding densities. First we need to establish that the assumptions in the beginning of Section 4 are satisfied.

The subexponential property implies that the tail the random variable X satisfies $\overline{F}_X(x) \sim n \overline{F}_Z(x)$, as $x \rightarrow \infty$. Hence, \overline{F}_X is regularly varying with index $-\alpha$

and the function U can be taken to be $n\overline{F}_Z$. Finally, we need to consider importance sampling algorithms with bounded relative error for computing rare event probabilities of the form $P(X > \lambda)$.

There exist several importance sampling algorithms for efficient computation of rare event probabilities of this form. Here we consider the dynamic mixture algorithms described in Hult and Svensson (2009) to generate N independent samples of X , denoted X_1, \dots, X_N . In particular we consider the conditional mixture algorithm of Dupuis et al. (2007) and the scaling mixture algorithm of Hult and Svensson (2009). Then, the tail e.d.f. $\overline{F}_{\nu_{u_p}, N}$ is constructed from the sample. Value-at-Risk is computed as $(1 - \overline{F}_{\nu_{u_p}, N})^\leftarrow(p)$ and expected shortfall as $\gamma_p((1 - \overline{F}_{\nu_{u_p}, N})^\leftarrow)$.

In the next subsection we verify the conditions of Proposition 3.4, Proposition 4.2, and Proposition 4.4 for these algorithms. Then the algorithms are implemented and their numerical performance is illustrated when Z_1 has a Pareto distribution.

5.1. Dynamic mixture algorithms. The dynamic mixture algorithm is designed for generating samples of X in (5.1) in order to efficiently compute rare event probabilities of the form $P(X > \lambda)$. Here it is convenient to use the notation $S_i = Z_1 + \dots + Z_i$, $i \geq 1$, $S_0 = 0$, and with this notation $X = S_n$ is the variable of interest. Each sample of S_n is generated sequentially by sampling Z_i from a mixture where the distribution of Z_i may depend on the current state, S_{i-1} . In the i th step, $i = 1, \dots, n-1$, where $S_{i-1} = s_{i-1}$, Z_i is sampled as follows.

- If $s_{i-1} > \lambda$, Z_i is sampled from the original density f_Z ,
- if $s_{i-1} \leq \lambda$, Z_i is sampled from

$$\begin{aligned} p_i f_Z(\cdot) + q_i g_i(\cdot \mid s_{i-1}), \quad & \text{for } 1 \leq i \leq n-1, \\ g_n(\cdot \mid s_{n-1}), \quad & \text{for } i = n, \end{aligned}$$

where $g_i(\cdot \mid s_{i-1})$ is a state dependent density. Here $p_i + q_i = 1$ and $p_i \in (0, 1)$.

The sampling measure distribution of S_n obtained by the dynamic mixture algorithm for computing $P(S_n > \lambda)$ is, throughout this section, denoted ν_λ .

The following results provide sufficient conditions for the upper bound $\varphi(c)$ that appears in Proposition 4.2 and Proposition 4.4.

Lemma 5.1. *Consider the mixture algorithm above with $p_i > 0$ for $1 \leq i \leq n-1$. Suppose there exist $a \in (0, 1)$ and $c > 0$ such that*

$$\liminf_{\lambda \rightarrow \infty} \inf_{\substack{s \leq c(1 - (1-a)^i) \\ y > a(c-s)}} \frac{g_i(\lambda y \mid \lambda s)}{f_Z(\lambda y)} \overline{F}_Z(\lambda) > 0, \quad 1 \leq i \leq n, \quad (5.2)$$

$$\limsup_{\lambda \rightarrow \infty} \sup_{\substack{s \leq c \\ y > c-s}} \frac{f_Z(\lambda y)}{g_n(\lambda y \mid \lambda s)} < \infty. \quad (5.3)$$

Then the scaled Radon-Nikodym derivative $\frac{1}{\overline{F}_Z(\lambda)} \frac{d\mu}{d\nu_\lambda}(\lambda y)$ is bounded on $\{y_1 + \dots + y_n > c\}$.

The proof is essentially identical to the proof of Lemma 3.1 in Hult and Svensson (2009) and therefore omitted.

Theorem 5.2. *Suppose (5.2) and (5.3) hold for $a \in (0, 1)$. Suppose, in addition, that there exist continuous functions $h_i : \mathbf{R}^n \rightarrow [0, \infty)$ and a constant $c_0 > 0$ such that*

$$\frac{f_Z(\lambda y_i)}{g_i(\lambda y_i \mid \lambda s_{i-1}) \overline{F}(\lambda)} \rightarrow h_i(y_i \mid s_{i-1}), \quad (5.4)$$

uniformly on $\{y \in \mathbf{R}^n : s_{i-1} \leq c(1 - (1 - a)^{i-1}), y_i > a(c - s_{i-1})\}$ for any $c \geq c_0$. Then, for $c \geq c_0$,

$$\limsup_{\lambda \rightarrow \infty} \frac{E_{\nu_\lambda} w_\lambda(X)^2 I\{X > c\lambda\}}{\overline{F}_X(\lambda)^2} \leq \sum_{i=1}^n \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \int_c^\infty h_i(y_i \mid 0) \alpha y_i^{-\alpha-1} dy_i, \quad (5.5)$$

with $q_n = 1$.

The proof is essentially identical to the proof of Theorem 3.2 in Hult and Svensson (2009) and therefore omitted.

By Theorem 5.2 we see that the function φ in Proposition 4.2 can be taken as the right-hand side of (5.5). Moreover, by Karamata's theorem, it is regularly varying with index $-\alpha$ if $h_i(y_i \mid 0)$ is slowly varying.

Finally, we establish that the conditions on the covariance function ϱ in Proposition 3.4 are satisfied.

Lemma 5.3. *Let X have distribution μ , d.f. F_X and density f_X . Suppose \overline{F}_X is regularly varying with index $-\alpha$, with $\alpha > 2$. Let ν denote any sampling distribution. If $d\mu/d\nu$ is bounded on (a, ∞) then the covariance function ϱ in (3.2) satisfies $\int_a^\infty \int_a^\infty \varrho(x, y) dx dy < \infty$ and $\varrho(x, x) = o([f_X(x)/\overline{F}_X(x)]^2)$.*

Proof. First note that if $d\mu/d\nu \leq K$ for some constant $K \in (0, \infty)$ then

$$\varrho(x, y) \leq K \overline{F}_X(y) - \overline{F}_X(x) \overline{F}_X(y), \quad \text{for } y \geq x.$$

Then

$$\begin{aligned} \int_a^\infty \int_a^\infty \varrho(x, y) dx dy &= 2 \int_a^\infty \int_x^\infty \varrho(x, y) dy dx \\ &\leq 2K \int_a^\infty \int_x^\infty \overline{F}_X(y) dy dx - 2 \left(\int_a^\infty \overline{F}_X(y) dy \right)^2. \end{aligned}$$

The first integral is finite, by Karamata's theorem, since $\alpha > 2$ and the second integral is finite for $\alpha > 1$ and then also for $\alpha > 2$. For the second condition

$$\frac{\varrho(x, x)}{f_X(x)^2 / \overline{F}_X(x)^2} \leq K \frac{\overline{F}_X(x)^3}{f_X(x)^2} - \frac{\overline{F}_X(x)^4}{f_X(x)^2}.$$

By Karamata's theorem $\alpha \overline{F}_X(x) \sim x f_X(x)$ so the expression in the last display is asymptotically equivalent to

$$K x^3 f_X(x) - x^4 f_X(x).$$

This converges to 0 as $x \rightarrow \infty$ when $\alpha > 2$ since f_X is regularly varying with index $-\alpha - 1$. This completes the proof. \square

5.2. Conditional mixture algorithms. The conditional mixture algorithm by Dupuis et al. (2007) can be treated with the above results.

The conditional mixture algorithm has, with $a \in (0, 1)$,

$$g_i(x | s) = \frac{f_Z(x)I\{x > a(b-s)\}}{\overline{F}_Z(a(b-s))}, \quad 1 \leq i \leq n-1,$$

$$g_n(x | s) = \frac{f_Z(x)I\{x > b-s\}}{\overline{F}_Z(b-s)}.$$

Then the techniques for establishing the conditions of Lemma 5.1 and Theorem 5.2, with $c_0 = 1$, are completely similar to the ones in Section 4.1 in Hult and Svensson (2009). The upper bound in Theorem 5.2 holds where the functions h_i are given by (see Hult and Svensson (2009))

$$h_i(y | s) = \lim_{\lambda \rightarrow \infty} \frac{f_Z(by)}{f_Z(\lambda y) / \overline{F}_Z(a\lambda(1-s)) \overline{F}_Z(\lambda) I\{y > a(1-s)\}}$$

$$= a^{-\alpha}(1-s)^{-\alpha}, \quad i = 1, \dots, n-1$$

and

$$h_n(y | s) = \lim_{\lambda \rightarrow \infty} \frac{f_Z(\lambda y)}{f_Z(\lambda y) / \overline{F}_Z(\lambda(1-s)) \overline{F}_Z(\lambda) I\{y > (1-s)\}} = (1-s)^{-\alpha}.$$

The resulting upper bound $\varphi(c)$ in Propositions 4.2 and 4.4 is given by

$$\varphi(c) = c^{-\alpha} \left(a^{-\alpha} \sum_{i=1}^{n-1} \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} + \prod_{j=1}^{n-1} \frac{1}{p_j} \right),$$

where $a \in (0, 1)$.

5.3. Scaling mixture algorithms. In the scaling mixture algorithm the large variables are generated by sampling from the original density and multiplying with a large number. In the context of scaling mixtures we assume that the original density f_Z is strictly positive on $(0, \infty)$. We also assume that $f_Z(x) = x^{-\alpha-1}L(x)$ with L slowly varying and $\inf_{x > x_0} L(x) =: L_* > 0$ for some $x_0 > 0$. The scaling mixture algorithm, with $\sigma > 0$, has

$$g_i(x | s) = (\sigma\lambda)^{-1} f_Z(x/\sigma\lambda) I\{x > 0\} + f_Z(x) I\{x \leq 0\}, \quad i = 1, \dots, n-1,$$

$$g_n(x | s) = (\sigma\lambda)^{-1} f_Z(x/\sigma\lambda) I\{x > 0, s \leq \lambda - \lambda(1-a)^{n-1}\}$$

$$+ f_Z(x) I\{x \leq 0 \text{ or } s > \lambda - \lambda(1-a)^{n-1}\}.$$

To generate a sample Z from g_i proceed as follows. Generate a candidate Z' from f_Z . If $Z' \leq 0$ put $Z = Z'$ and if $Z' > 0$, put $Z = \sigma\lambda Z'$.

For the scaling mixture algorithm the conditions of Lemma 5.1 and Theorem 5.2 can be established with $c_0 < 1$. The techniques for doing this are completely similar to the ones in Section 4.3 in Hult and Svensson (2009). The upper bound in Theorem 5.2 holds where the functions h_i are given by (see Hult and Svensson (2009))

$$h_i(y_i | s_{i-1}) = \alpha \lambda [y_i^{\alpha+1} f(y_i/\lambda)]^{-1}.$$

and the resulting upper bound $\varphi(c)$ in Propositions 4.2 and 4.4 is given by

$$\varphi(c) = \sum_{i=1}^n \prod_{j=1}^{i-1} \frac{1}{p_j} \frac{1}{q_i} \int_c^\infty \frac{\alpha}{\lambda^\alpha L(y_i/\lambda)} \alpha y_i^{-\alpha-1} dy_i.$$

5.4. Numerical computation of Value-at-risk. We now consider a sum of n Pareto-distributed random variables,

$$S_n = Z_1 + \dots + Z_n.$$

We will estimate quantiles of S_n by using the importance sampling e.d.f. given by the scaling mixture algorithm in Hult and Svensson (2009) (SM) as well as the conditional mixture algorithm in Dupuis et al. (2007) (DLW). The changes of measure are chosen by using the asymptotic approximation of the quantiles,

$$\lambda_p^* = (n/(1-p))^{1/\alpha} - 1.$$

This approximation is based on the subexponential property, and since $P(S_n > x) > nP(X_1 > x)$ for positive random variables, it is smaller than the true quantile.

For p equal to 0.99, 0.999 and 0.99999, we use the DLW algorithm 10^2 times with $N = 5 \cdot 10^4$ samples to obtain a reference value which we refer to as the true value of the quantile.

We compare the performance of the quantile estimates based on $N = 10^4$ samples. The estimation is repeated 100 times and the mean and standard deviation of the estimates are reported.

We also include the results from standard Monte Carlo for comparison.

TABLE 1. Simulations of λ_p such that $P(S_n > \lambda_p) = 1 - p$, where $S_n = \sum_{i=1}^n Z_i$ and $P(Z_1 > x) = (1+x)^{-2}$. The number of samples used for each estimate was $N = 10^4$ and the estimation was repeated 100 times.

n	$1-p$	True	Approx.	SM	DLW	MC	
10	1e-2	40.141	30.623	41.007 (0.246)	40.166 (0.459)	40.038 (1.780)	Avg. est. (Std. dev.)
	1e-3	108.49	99.000	109.33 (0.847)	108.29 (1.081)	84.821 (47.23)	
	1e-5	1007.4	999.00	1003.1 (18.5)	1007.5 (1.51)	609.42 (1594)	
30	1e-2	84.622	53.772	85.841 (0.3950)	84.681 (1.237)	84.362 (2.739)	
	1e-3	202.41	172.21	203.56 (1.530)	202.29 (2.400)	171.16 (71.26)	
	1e-5	1759.5	1731.1	1753.7 (41.12)	1759.0 (1.487)	114.23 (443.5)	

5.5. Numerical computation of expected shortfall. Using the setting from the previous section, we also calculate the expected shortfall for the case of a random walk with Pareto-distributed increments. We first consider the case where $\alpha = 2$, although it does not satisfy the conditions of Proposition 4.4.

TABLE 2. Simulations of λ_p such that $P(S_n > \lambda_p) = 1 - p$, where $S_n = \sum_{i=1}^n Z_i$ and $P(Z_1 > x) = (1 + x)^{-3}$. The number of samples used for each estimate was $N = 10^4$ and the estimation was repeated 100 times.

n	$1 - p$	True	Approx.	SM	DLW	MC	
10	1e-2	14.190	9.0000	14.853 (0.090)	14.195 (0.154)	14.182 (0.305)	Avg. est. (Std. dev.)
	1e-3	25.656	20.544	26.125 (0.171)	25.588 (0.412)	24.965 (2.212)	
	1e-5	103.42	99.000	104.23 (0.799)	103.40 (0.553)	5.283 (16.03)	
30	1e-2	29.951	13.422	31.054 (0.287)	29.943 (0.519)	29.949 (0.500)	
	1e-3	46.072	30.072	46.725 (0.286)	46.277 (1.041)	44.608 (2.688)	
	1e-5	157.65	143.22	158.46 (1.080)	157.62 (0.273)	13.847 (28.53)	

TABLE 3. Simulations of $E(S_n | S_n > \lambda_p)$, where $P(S_n > \lambda_p) = 1 - p$, $S_n = \sum_{i=1}^n Z_i$ and $P(Z_1 > x) = (1 + x)^{-2}$. The number of samples used for each estimate was $N = 10^4$ and the estimation was repeated 100 times.

n	$1 - p$	True value	SM	DLW	MC	
10	1e-2	71.795	73.065 (1.06) [0.845]	71.831 (1.22) [0.815]	72.252 (8.75) [0.702]	Avg. est. (Std. dev.) [Avg. time (s)]
	1e-3	208.84	209.37 (3.60) [0.734]	209.30 (4.99) [0.724]	213.42 (65.8) [0.572]	
	1e-5	2008.4	2009.8 (37.1) [0.866]	2009.3 (30.9) [0.822]	4787.8 (23168) [0.693]	
30	1e-2	139.22	140.55 (2.22) [1.189]	139.14 (3.09) [1.077]	140.76 (17.34) [0.903]	
	1e-3	376.29	375.76 (5.00) [1.033]	378.24 (11.49) [0.936]	391.06 (96.36) [0.757]	
	1e-5	3494.4	3500.0 (65.2) [1.301]	3496.9 (59.8) [1.224]	745.70 (3671) [0.991]	

REFERENCES

- Bassamboo, A., Juneja, S. and Zeevi, A., 2007. On the efficiency loss of state-independent importance sampling in the presence of heavy-tails. Oper. Res. Lett.

TABLE 4. Simulations of $E(S_n|S_n > \lambda_p)$, where $P(S_n > \lambda_p) = 1 - p$, $S_n = \sum_{i=1}^n Z_i$ and $P(Z_1 > x) = (1 + x)^{-3}$. The number of samples used for each estimate was $N = 10^4$ and the estimation was repeated 100 times.

n	$1 - p$	True value	SM	DLW	MC	
10	1e-2	19.260	20.044 (0.167) [0.702]	19.257 (0.395) [0.727]	19.495 (0.905) [0.605]	Avg. est. (Std. dev.) [Avg. time (s)]
	1e-3	36.658	36.911 (0.327) [0.7079]	36.463 (0.776) [0.731]	41.032 (5.53) [0.613]	
	1e-5	154.74	154.39 (1.326) [0.708]	153.83 (2.705) [0.733]	132.74 (491.7) [0.607]	
30	1e-2	37.277	38.603 (0.902) [0.885]	37.200 (1.169) [0.923]	37.744 (1.581) [0.707]	
	1e-3	62.090	62.013 (0.416) [0.912]	62.066 (1.814) [0.939]	69.369 (7.973) [0.712]	
	1e-5	232.01	230.27 (1.92) [0.911]	230.00 (1.47) [0.935]	225.14 (932) [0.703]	

35, 251-260.

- Blanchet, J. and Liu, J., 2008. State dependent importance sampling for regularly varying random walks. *Adv. Appl. Probab.* 40, 1104-1128.
- Blanchet, J. and Glynn, P., 2008. Efficient rare event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Probab.* 18, 1351-1378.
- Csörgő, M., Csörgő, S., and Horváth, L., 1986. An asymptotic theory for empirical reliability and concentration processes. *Lecture Notes in Statistics*, 33. Springer, Berlin.
- Dupuis, P., Leder, K. and Wang, H., 2007. Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* 17(3).
- Glasserman, P., Heidelberger, P. and Shahabuddin, P., 2002. Portfolio Value-at-Risk with heavy-tailed risk factors. *Math. Financ.* 12, 239-270.
- Glynn, P., 1996. Importance sampling for Monte Carlo estimation of quantiles. *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation* 180-185. Publishing House of Saint Petersburg University.
- Hult, H. and Svensson, J., 2009. On importance sampling with mixtures for random walks with heavy tails. Preprint, KTH.
- van der Vaart, A.W. and Wellner, J.A., 1996. *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

(H. Hult) DEPARTMENT OF MATHEMATICS, KTH, 100 44 STOCKHOLM, SWEDEN
E-mail address: `hult@kth.se`

(J. Svensson) DEPARTMENT OF MATHEMATICS, KTH, 100 44 STOCKHOLM, SWEDEN
E-mail address: `jenssve@kth.se`